Materials-Based 3D Segmentation of Unknown Objects from Dual-Energy Computed Tomography Imagery in Baggage Security Screening

Andre Mouton[a,*], Toby P. Breckon[b]

[a] *School of Engineering, Cranfield University, U.K.*
[b] *School of Engineering and Computing Sciences, Durham University, U.K.*

Abstract

We present a novel technique for the 3D segmentation of unknown objects from cluttered dual-energy Computed Tomography (CT) data obtained in the baggage security-screening domain. Initial materials-based coarse segmentations, generated using the Dual-Energy Index (DEI), are refined by partitioning at automatically-detected regions. Partitioning is guided by a novel random forest based quality metric, trained to recognise high-quality, single-object segments. A second novel segmentation quality measure is presented for quantifying the quality of full segmentations based on the random forest metric of the constituent parts and the error in the number of objects segmented. In a comparative evaluation between the proposed approach and three state-of-the-art volumetric segmentation techniques designed for single-energy CT data (two region-growing [1, 2] and one graph-based [3]) our method is shown to outperform both region-growing methods in terms of segmentation quality and speed. Although the graph-based approach generates more accurate partitions, it is characterised by high processing times and is significantly outperformed by the proposed method in this regard. The observations made in this study indicate that the proposed segmentation technique is well-suited to the baggage security-screening domain, where the demand for computational efficiency is paramount to maximise throughput.

Keywords: Segmentation, Dual-energy computed tomography, Random forests, Baggage-CT imagery

1. Introduction

The central role of baggage screening in the aviation-security domain has lead to an increased interest in the development of image-processing and computer-vision techniques to

*Principal Corresponding Author
Email address: andremouton.email@gmail.com (Andre Mouton)

advance the automated analysis of baggage imagery. 3D X-ray Computed Tomography (CT), which has enjoyed much success in medical imaging, has fairly recently been introduced to the security-screening domain in an attempt to mitigate the limitations of conventional 2D X-ray imagery (e.g. occlusion, clutter and density confusion) [4, 5]. Material-based discrimination is thus possible using the correlations between the effective atomic numbers and densities of materials and has formed the basis of automated explosives detection in security-screening applications.

The foremost application of CT within the security-screening domain has been the materials-based detection of explosives [6]. Dual-Energy Computed Tomography (DECT) [7], whereby objects are scanned at two distinct energies, provides an effective means for performing such materials-based discrimination. Owing to this primary explosives detection-based objective of imaging within the aviation-security domain, DECT machines have been the baggage-CT scanners of choice. The primary, non-object recognition-based objective of typical dual-energy baggage-CT scanners, however, coupled with the demand for high throughput, means that 3D baggage-CT imagery typically presents with substantial noise, metal-streaking artefacts and poor voxel resolution and is thus generally of a poorer quality than medical CT imagery [6] (Figure 1). Recent work in this domain has been concerned primarily with the automated classification of illicit materials in baggage-CT imagery [8–10]. Although promising results have been presented, these have been achieved on manually segmented subvolumes. The automated segmentation of low-resolution, cluttered volumetric baggage-CT imagery remains an open problem.

While image segmentation is a core problem in computer vision and has been the source of an extensive resource of literature, prior work addressing the automated segmentation of unknown objects from low-resolution, cluttered volumetric baggage-CT imagery in particular is limited [1, 11, 12]. The differences in the quality and the nature of security-screening imagery compared to medical imagery, have limited the success of medical-segmentation techniques in this domain [13]. Furthermore, the dependence of the majority of the state-of-the-art medical-segmentation techniques on *a priori* information in particular, detracts from their suitability to the security-screening imagery, where the segmentation of multiple, unknown objects is required. This has been substantiated by the comparative study of Megherbi *el al.* [13], where popular medical-segmentation techniques are shown to perform poorly on baggage-CT data. A recent study [14] presents a novel technique, using curvature estimation of 2D contours, for the automated segmentation of knots in 3D CT images of wet wood. Although

the study presents state-of-the-art results (both in terms of accuracy and speed), the contour estimation exploits a geometric prior and therefore, similarly to the aforementioned medical segmentation techniques, is ill-suited to baggage CT segmentation.

Unsupervised (blind) segmentation of unknown objects from cluttered volumetric imagery is considerably more challenging and comparatively few solutions exist. The majority of the prior work in this domain has originated from the US Department of Homeland Security's Awareness and Localization of Explosives-Related Threats (ALERT) initiative [11] which led to the development of five dedicated baggage-CT segmentation techniques. Wiley *et al.* [1] present a 3D flood-fill region-growing method based on the Stratovan Tumbler medical-segmentation technology [15]. Although high-quality segmentations are presented, performance is shown to rely on near isotropic voxel resolutions in all three dimensions. Song *et al.* [11] present a sequential *segment-and-carve* approach using the seedless (i.e. unsupervised) Symmetric Region-Growing (SymRG) technique [2] coupled with extensive post-segmentation processing (splitting and merging). The approach is extremely convoluted (characterised by a large parameter set) and optimal performance is shown to rely on careful parameter tuning. Grady *et al.* [12] perform unsupervised segmentation of baggage-CT data via recursive partitioning using the linear-time isoperimetric distance-tree algorithm [3]. Segmentation is guided by a novel Automated Quality Assessment (AQUA) measure, which automatically computes the quality of a given segmentation without *a priori* knowledge of the object being segmented. High-quality segmentations are achieved at relatively low processing times. Harvey *et al.* [11], present a multi-scale baggage-CT segmentation approach using the sieves class of algorithms [16] and Sedgewick's connected component analysis technique [17], while Feng *et al.* [11] present an approach which appears to be largely based on the region-growing and merging technique of Ugarriza *et al.* [18]. While the approach is shown to yield high-quality segmentations, satisfactory results are again reliant on careful parameter tuning.

It is worth emphasising that each of the five aforementioned baggage-CT segmentation techniques were developed and evaluated using a fully labelled volumetric baggage-CT data set captured on a single-energy medical-grade CT scanner with a near isotropic voxel resolution of $0.98 \times 0.98 \times 1.29$mm. Furthermore, the data was entirely free of threats/contraband and contained comparatively low numbers of metallic objects. Such data is not representative of the current benchmark in baggage screening [8, 10, 19]. The segmentation of low, anisotropic resolution volumetric baggage-CT imagery (obtained on dual-energy scanners) in the presence of multiple metal objects has not been considered previously.

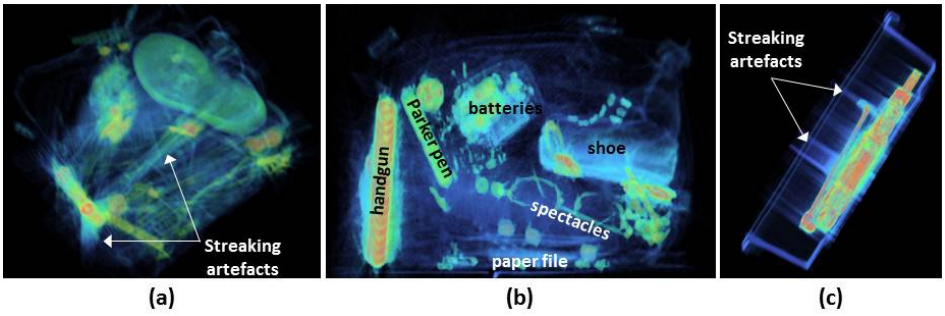

Figure 1: Baggage-CT scans illustrating poor image quality, low resolution, artefacts and clutter (obtained on Reveal CT80-DR dual-energy baggage scanner).

Here, we address these limitations by presenting a dual-energy-based segmentation technique which is shown to produce fast, high-quality segmentations of low-resolution volumetric baggage-CT images. Our approach is based on four novel contributions: 1) a materials-based coarse segmentation technique using the Dual-Energy Index (DEI) [7] (Section 2) and connected component analysis; 2) a random-forest-based model for measuring the quality of individual object segments, which is used to guide the segmentation process; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) an efficient segmentation-refinement procedure for splitting fused objects. The current state-of-the-art in baggage-CT-image classification [8, 20] relies on the manual generation of subvolumes containing at most a single target object (akin to the use of a sliding window in traditional image-based object recognition). The objective of the segmentation techniques developed in this study thus seek to automate this procedure by segmenting every object in a given scan and generating a corresponding subvolume for each.

We thus extend upon the prior literature [1, 11, 12] by considering the segmentation of low-resolution, cluttered baggage-CT imagery obtained on a dual-energy CT scanner, which is representative of that encountered within an aviation security environment [19, 21].

2. Dual-Energy Computed Tomography (DECT)

Conventional, single-energy CT systems produce reconstructions representative of the Linear Attenuation Coefficients (LAC) of the object under investigation [22]. That is to say, the greyscale intensity values (i.e. CT numbers, in Hounsfield Units (HU)) in the CT image are dependent on the LAC of the scanned object. Consequently, it becomes challenging and in some cases, impossible, to distinguish between materials that share similar LACs. In contrast, Dual-Energy CT (DECT) techniques, whereby attenuation data is captured using two distinct X-ray spectra, offer a means for characterising the chemical composition (e.g. atomic

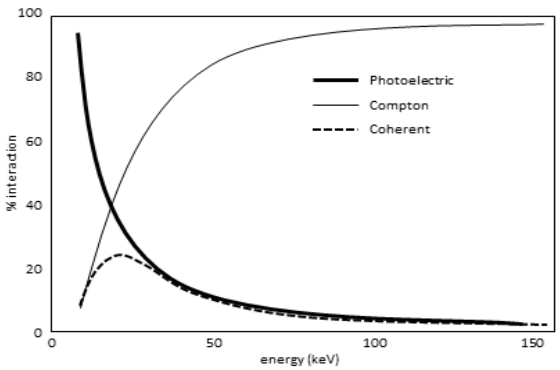

Figure 2: Relative contribution of different modes of interactions as a function of energy (for water) [22].

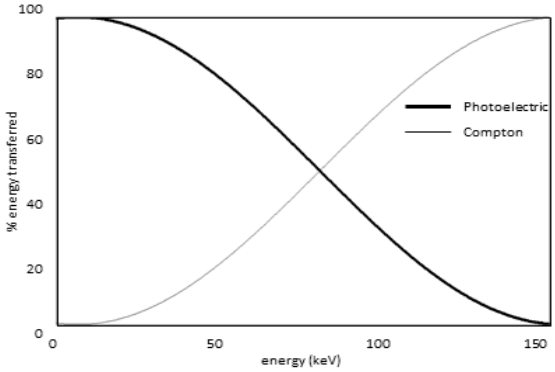

Figure 3: Energy transfer percentage of the different interactions (in water) [22].

number and/or density) of the material under investigation based on its response under these different spectral conditions.

The physical basis of DECT imaging relies on the energy dependence of the interaction of X-ray photons with matter (Figures 2 and 3). More particularly, within a photon energy range of approximately 30 keV to 200 keV, these interactions are known to be dominated by the photoelectric effect and Compton scattering [23]. Dual-energy CT involves acquiring attenuation measurements for an object at two different tube voltages (usually 80 kVp and 140 kVp for medical applications), resulting in two separate attenuation profiles. The dual-energy decomposition problem is then to determine the Compton scatter and the photoelectric absorption coefficients of the material from the measured high and low-energy projections. Alternatively, it has been shown that the attenuation coefficients for any material may be expressed as a linear combination of the coefficients of two basis materials, provided that the two chosen materials are sufficiently different in their atomic numbers (and hence in their Compton and photoelectric coefficients) [24]. The decomposed dual-energy data may then be used to compute the effective atomic numbers and electron densities of the materials present in a scan [25] (this is the basis for traditional explosives detection systems in aviation security-screening [6]).

DECT techniques typically fall into one of three categories [26]: 1) post-reconstruction techniques; 2) pre-reconstruction techniques and 3) iterative-reconstruction techniques. Pre-reconstruction [27] and iterative reconstruction [28] techniques are generally concerned with solving the material decomposition problem and subsequently determining the effective atomic number and density images. While a variety of DECT-decomposition techniques exist (e.g. direct approximation [23]; iso-transmission lines [29]), they typically require access to raw projection data and rely on physical calibration procedures [27].

A more rudimentary estimation of the chemical characteristics of a scan may be obtained via post-reconstruction (or image-based) DECT techniques. In contrast to pre-reconstruction and iterative reconstruction DECT, these do not require a calibration procedure or the availability of raw-data. The Dual Energy Index (DEI) [7] is a post-reconstruction DECT measure that offers a crude estimate of the chemical characteristics of a scan. For a material in air (i.e. not dissolved in water) the DEI is given by [7]:

$$DEI = \frac{x_L - x_H}{x_L + x_H + 2000} \tag{1}$$

where x_L and x_H are the pixel values, in Hounsfield Units (HU), for the low and high-energy scans respectively. While the DEI is an indicator of the effective atomic number of a material, in contrast to the true effective atomic number [25], its value does not rely on the photoelectric cross-section characteristics of the material (which are not precisely known) [7]. Despite its ease of computation, the DEI has demonstrated potential in material differentiation for a variety of clinical tasks, including: the differentiation of air and tagged faecal materials from soft-tissue colonic structures in CT colonography [30] and the chemical characterisation of urinary stones in abdominal CT imagery [31].

The predominant limitation of post-reconstruction DECT is its susceptibility to artefacts in the reconstructed images [28]. Therefore, although the DEI has been successfully applied to a variety of clinical tasks [7, 31–33], it has not previously been considered in the baggage-CT domain.

3. Proposed Segmentation Approach

The refinement of initial coarse segmentations is a popular approach in both the medical [34–36] and security-screening [12] segmentation literature. We consider a similar approach, composed of the following three components: 1) coarse segmentation/foreground determination; 2) segmentation refinement; 3) segmentation quality measurement.

3.1. Coarse Segmentation

Low and high-energy volumes are pre-processed with a Metal Artefact Reduction (MAR) technique [37]. The artefact-reduced images are used to compute the Dual-Energy Index (DEI) image according to Equation 1 [7] (Section 2).

Based on the assumption that different materials are relatively well separated by their DEI, a set of N_τ material-specific images is generated by subjecting the DEI image to a multiple-thresholding procedure:

$$I_{\tau_i} = \begin{cases} 1 & \text{if } \tau_{i-1} \leq I_{\text{dei}} \leq \tau_i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \ldots, N_\tau \tag{2}$$

where the number of thresholds N_τ is a user specified parameter. Connected component analysis [38] is performed on each of the N_τ thresholded images. All connected components smaller than a predefined minimum object size are discarded while the remaining connected components are assigned individual labels and represent the image foreground (or coarse segmentation) to be passed to the refinement algorithm. While it is likely that an improvement in the quality of the coarse segmentations will result from a case-by-case fine-tuning of the DEI thresholds (e.g. by DEI histogram analysis), to ensure automation and maintain low processing times, a predefined, constant set of thresholds is used. To this end, a constant uniformly-space threshold range is chosen to optimise processing times and to ensure that consistent material types are segmented across all images. Further segmentation, or refinement (driven by a segmentation-quality metric), of each of the coarsely segmented components in an image is used to generate the final segmented image.

3.2. Segmentation Quality Measure

The segmentation-quality measure is intended to provide a quantification of the likelihood that a given segment represents a single object (i.e. does not require further segmentation). For the purposes of online segmentation evaluation (i.e. evaluation in the absence of ground-truth), the feature-based generative model of high-quality segmentations presented by Kohlberger *et al.* [39] has been met with success in related studies [12]. More specifically, Grady *et al.* [12] present the Automated QUality Assessment (AQUA) measure to quantify segmentation quality and control the splitting of connected components within their graph-partitioning approach. Using a variety of segmentation metrics as features [39], the AQUA module is trained (using a Gaussian Mixture Model (GMM)) on a training set of single-object features to recognise high-quality (i.e. high likelihood of consisting of only a single object) segmentations. The model is shown to provide meaningful separation of good (or single) and bad (or multi) object segments [11].

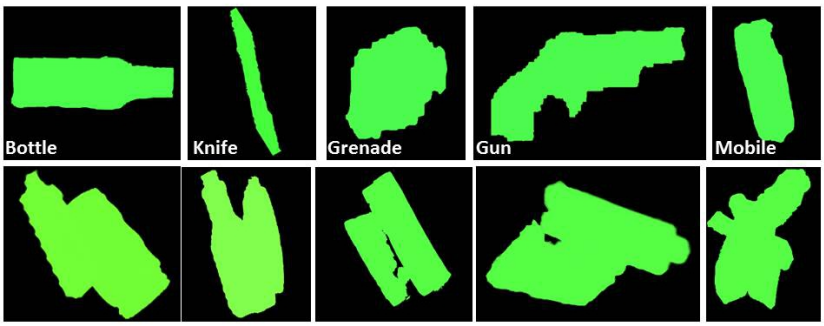

Figure 4: Object segment mask examples. Top row: manually generated single-object segments. Bottom row: manually generated multi-object segments.

We examine the suitability of the AQUA measure, using the current dataset, and the 42D feature vector described in [12, 39]. The feature vector quantifies the shape and appearance of object segments by computing 42 feature attributes falling into one of five categories: 1) unweighted geometric features (quantifying the size and regularity of the segmentation); 2) weighted geometric features (locally emphasising the geometric features when intensity values are similar to each other); 3) intensity features (measuring absolute intensity and intensity distributions within segmentations); 4) gradient features and 5) ratio features (computed as ratios of previously computed features). For a detailed description of the features used, the reader is referred to [39] and [12]. Similarly to [12], the features are extracted from a set of manually segmented single-object segments. A large range of objects are included in this training set (Figure 4). Principal Component Analysis (PCA) [40] is applied to reduce the dimensionality of the feature space. The GMM is fitted over the PCA coefficients of all the segments in the training set using the Expectation-Maximisation (EM) technique [41]. In determining the optimal GMM, the number of Gaussians and the covariance matrix type (diagonal or full-rank) are varied and the model minimising the Bayesian Information Criterion (BIC) [42] cost function is selected. The AQUA measure of a given object segment $S \subseteq V$ (where V is the volume) is then given by [12]:

$$\mathrm{AQUA}(S) = \sum_{i=1}^{10} w_i \mathcal{N}(f(S); \mu_i, \Sigma_i) \tag{3}$$

Although the motivation for using a generative model for discriminating between single and multi-object segments is based on the premise that multi-object segments exhibit greater variability in feature space than single-object segments [12], the most significant contributions in the related field of multi-organ segmentation in 3D CT imagery have been achieved using random forest-based classifiers [43–45]. In fact, random forests have enjoyed increasing

popularity in complex medical-image segmentation tasks in general and have been successfully applied to the segmentation of adrenal gland abnormalities in CT imagery [46]; synaptic contacts in electron microscopy images [47]; foetal brain structures in ultrasound images [48]; the myocardium in real-time 3D echocardiography [49] and a range of structures in MR imagery (e.g. multiple sclerosis [50, 51]; high-grade gliomas [52]; left ventricle [53], neurological structures [54, 55]). In accordance with these trends in the prior literature, we thus propose investigating the efficacy of a novel random-forest-based discriminative model to distinguish between good and bad object segments. In addition to the aforementioned 42D descriptor (hereafter denoted *Kohlberger*) [12, 39] and as an additional novel aspect within this work, the performance of three shape-based descriptors are considered as it is expected that the characteristics of the surface of a single-object segment will differ from that of a segment representing multiple objects. 3D shape-based descriptors have been successfully applied to a variety of similar object-recognition, retrieval and classification tasks [56–58]. Based on these prior works, the following three 3D shape-based descriptors are considered: 1) 3D Zernike descriptors (denoted *Zernike*) [59]; 2) the Histogram-of-Shape Index (denoted *HSI*) [60] and 3) a hybrid 3D shape descriptor (denoted *Shape*) [58].

Novotni and Klein [59] developed the 3D Zernike descriptors by expanding upon the mathematical concepts of 3D Zernike moments as laid out by Canterakis [61]. 3D Zernike polynomials are a set of basis functions that are orthogonal on the unit sphere. The 3D Zernike descriptor is an extension of the spherical harmonics-based descriptors of Kazhdan *et al.* [62] and have been shown to be compact, robust to noise and invariant to rotation [59]. By combining a set of radial basis functions with spherical harmonics, the n^{th}-order 3D Zernike functions may be computed as [61]:

$$Z_{nl}^{m}(r,\theta,\phi) = R_{nl}(r)Y_{l}^{m}(\theta,\phi) \tag{4}$$

where Y_l^m are complex valued spherical harmonics defined on the spherical coordinate system given by (θ, ϕ); n, l, m are integers such that $|m| \leq n$ and $n - |m|$ is even and $R_{nl}(r)$ are orthogonal radial basis polynomials [63]. Equation 4 may be rewritten in Cartesian coordinates:

$$Z_{nl}^{m}(\mathbf{x}) = \sum_{v=0}^{k} q_{kl}^{v} |\mathbf{x}|^{2v} e_{l}^{m}(\mathbf{x}) \tag{5}$$

where $2k = n - l$ and $e_l^m = r^l Y_l^m(\theta, \phi)$ are the harmonic polynomials as defined by [62]. The coefficients q_{kl}^v are computed such that the functions are orthonormal in the unit sphere [59]. The Zernike moments are then the projection of a given 3D shape function onto this orthonormal basis:

$$\Omega_{nl}^m = \frac{3}{4\pi} \int_{||x|| \leq 1} f(\mathbf{x}) \overline{Z_{nl}^m(\mathbf{x})} d\mathbf{x} \tag{6}$$

where the voxelised 3D shape function $f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^3$ is a binarised representation of the object surface, defined on a regular cubic grid. While these moments are *not* invariant under rotations, Novotni and Klein [59] achieved rotational invariance by collecting the moments into $(2l+1)$-dimensional vectors: $\Omega_{nl} = (\Omega_{nl}^l, \Omega_{nl}^{l-1}, \ldots, \Omega_{nl}^{-l})^t$. The rotationally-invariant, 3D Zernike descriptors F_{nl} are subsequently defined as the norms of the vectors Ω_{nl} [59]:

$$F_{nl} = ||\Omega_{nl}||_2 = \sqrt{\sum_{m=-l}^{l} (\Omega_{nl}^m)^2} \tag{7}$$

Dorai and Jain [60] present the Histogram-of-Shape Index (HSI) for the representation and recognition of arbitrarily curved rigid 3D objects. The Shape Index (SI) is a scalar-valued quantitative measure of the shape of a surface at a point p:

$$\mathrm{SI}(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{8}$$

where κ_1 and κ_2 ($\kappa_1 \geq \kappa_2$) are the principal curvatures of the surface at the point p [60]. Given a smooth, plane unit-speed (parametrised) curve $\gamma(t)$, where t is the arc-length, the curvature is defined as:

$$K(t) = ||\ddot{\gamma}|| \ := \frac{1}{r} \tag{9}$$

where r is the radius of the osculating (kissing) circle. If M is defined as differentiable surface in $\mathbb{R}^3$, then at each point p on M, the surface has two principal curvatures κ_1 and κ_2, computed as follows [64].

Let $\mathbf{n}$ be the unit normal vector to M at p and determine the normal plane $P \in \mathbb{R}^3$ containing $\mathbf{n}$. The unit-speed curve $\gamma_P(t) = P \cap M$ (i.e. the plane curve resulting from the intersection of P and M) is then determined. Subsequently, the curvature K_P of $\gamma_P(t)$ is computed according to Equation 9 and finally, the principal curvatures are defined as $(\kappa_1, \kappa_2) = (\min_P K_P, \max_P K_P)$. According to this definition, every distinct shape may be mapped onto a unique value in the interval SI $\in [0, 1]$. The exception to this is the planar shape, for which SI is undefined since $\kappa_1 = \kappa_2 = 0$ for all points on a planar surface [60].

Megherbi *et al.* [58] propose combining the HSI and Zernike descriptors (by direct concatenation) yielding a hybrid 3D shape descriptor. The proposed descriptor demonstrates potential in the classification of threats in CT-baggage imagery and is included in the evaluation here. Prior to the extraction of the aforementioned shape-based features, the object segments are pose-normalised to ensure invariance to changes in scale and translation. This is achieved by translating and rescaling (voxel resampling by nearest-neighbour interpolation) each object segment based on its approximate minimum-bounding box within the original CT image [58].

For a given descriptor type, individual trees in the random forest are constructed in a top-down recursive manner using a simple thresholding function as the node split function for all internal nodes. The optimality criterion used for node splitting is the classical Information Gain (IG) [65]. Randomness is injected into the trees via random node optimisation, whereby a random subset of the available node test parameter values is considered at each node. Trees are grown to a maximum depth D_T and leaf nodes are generated if the IG falls below a minimum threshold $IG_{\min}$. The quality of any given segmentation - denoted the Random Forest Score (RFS) - is computed by averaging the corresponding posterior probabilities of each of the leaf nodes reached in the forest:

$$\text{RFS} = p(c|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^{T} p_t(c|\mathbf{v}) \tag{10}$$

where T is the number of trees in the forest; $p(c|\mathbf{v})$ is the estimated conditional probability that a given test point $\mathbf{v}$ belongs to the class c and c is a discrete class label (i.e. $(0, 1) \rightarrow$ (bad,good)).

3.3. Segmentation Refinement

Based on the results of preliminary experimentation (see Section 5.1) regarding the aforementioned segmentation-quality measures, a simple yet efficient technique for refining the

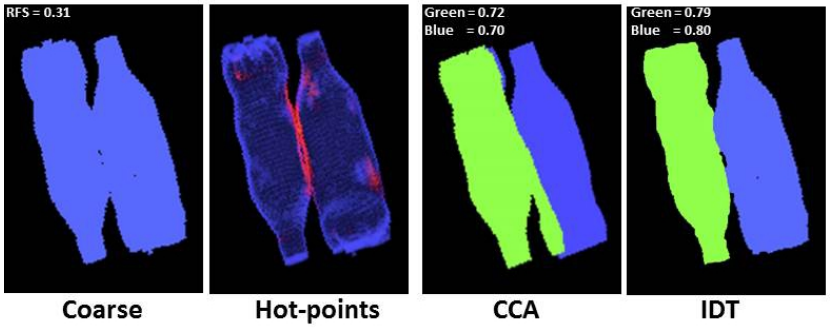

Figure 5: Segmentation refinement examples using CCA and IDT [12]. Hot-points (red) and RFS indicated.

initial, coarse segmentation using a random-forest-based approach is proposed. The RFS is computed for each of the N_c components (or objects) in a given coarse segmentation. Those components yielding an RFS below a given threshold τ_{RFS} are considered to be composed of multiple objects and are passed to a partitioning approach. Coarse components with RFS$> \tau_{RFS}$ are left unchanged and assigned a unique label in the final image. Poor-quality objects (RFS$< \tau_{RFS}$) are partitioned at the estimated intersection (or touching) points of the multiple objects comprising the given segment. These points are found by detecting the perimeter voxels of the original object that are likely to be common to two objects. Non-zero (i.e. object) voxels are assumed to lie on the perimeter of the object if they are connected to at least one zero (i.e. background) voxel. The assumption is made that those voxels corresponding to the intersections of multiple objects will be surrounded by higher numbers of object voxels compared to regular (non-intersection) perimeter points [11]. The total number of object voxels in a predefined local cubic neighbourhood ($11 \times 11 \times 11$) of each perimeter voxel is thus determined. If this number is greater than a predefined threshold, τ_{HP}, the perimeter voxel is considered to be an intersection point [11] (red points in Figure 5). For a given object, this analysis may result in multiple clusters of such points (denoted *hot-points*), in which case it is assumed that the object requires splitting at multiple regions. Each cluster of hot-points is considered individually. It has previously been suggested that splitting of touching objects may be performed by fitting a plane (e.g. by RANSAC [66]) to such hot-points [11]. Such planes, however, are likely to intersect the object at multiple regions (not just at the locations of the hot-points) leading to over-segmentations. Restricting the planes to local regions is challenging, especially when determining which voxels lie above or below the plane. The plane-based approach becomes particularly problematic when an object requires splitting at multiple locations.

A simpler approach is proposed here, whereby Connected Component Analysis (CCA) is

performed on a mask obtained by removal (i.e. setting to zero) of the hot-points. The two connected components returning the highest RFS (computed with the hot-points reinserted) are retained. If the CCA results in only a single connected component, morphological dilation of the zeroed-out region is performed until the CCA returns at least two components. In the case where the RFS of one of the regions falls below the RFS of the original region, or the region is smaller than the minimum permissible size of an object, the region is assumed to be noise or artefact-induced and is discarded. If both components result in a decrease relative to the original RFS, then the original object is retained (Figure 5). Although objects split in this way are not guaranteed to produce segments with RFS$> \tau_{RFS}$, only splits resulting in improved scores are permitted. The procedure thus performs both splitting as well as denoising of the coarse segmentations. For objects containing multiple hot-point clusters, the RFS of the final split objects are affected by the order in which the clusters are considered. As the described splitting procedure is fast and the numbers of hot-point clusters per object are generally low (typically ≤ 3), the optimal order (i.e. that which results in the individual objects with the highest RFS) may be determined by testing all possible orders.

4. Comparative Methodologies

The performance of the proposed segmentation approach (denoted *CCA*) is evaluated in comparison to the isoperimetric distance tree approach [3], symmetric region-growing [2] and 3D flood-fill region-growing [1]. A brief overview of each technique is presented here, for detailed explanations the reader is referred to the original works [1–3, 12, 67].

Isoperimetric Distance Trees (IDT) [3] are evaluated as an alternative to the proposed segmentation-refinement procedure. IDT has previously demonstrated success as part of an automated segmentation approach for medical-grade CT imagery [3, 12]. As opposed to applying the identical techniques used in [12], IDT is used as a direct alternative to the proposed segmentation-refinement method - that is to say, the technique is applied to the DEI coarse segmentations and is driven by the RFS (Equation 10). The IDT technique is applied recursively to each individual connected component in the DEI mask until one of three criteria is met [12]: 1) the input mask has a sufficiently high quality (as determined by the RFS); 2) the partitioned objects have RFS below a given threshold τ_γ and an isoperimetric ratio greater than that of the input mask (i.e. further partitioning produces low-quality segmentations) or 3) the input mask is smaller than the minimum permissible object size.

Symmetric Region Growing (SymRG) [2] is a seedless (parameter-free), single-pass

region-growing technique which has demonstrated success in a variety of 3D segmentation tasks [2, 68]. For any voxel $p \in R_i$ where $R_i \subset V$ is an evolving region and V is the input volume, a neighbouring voxel q will be included in the region R_i if the function $g(p, q)$ is true, where $g(.,.)$ is a symmetric function satisfying:

$$g(p, q) = g(q, p) \ \ \forall p, q \in V \tag{11}$$

The main advantage of SymRG is its invariance to the region start points (seed points) and its low processing times - it is not claimed to provide higher-quality segmentations compared to existing methods [2]. Nonetheless, its fully-automated nature and low processing times make SymRG an attractive option in the security-screening domain.

3D Flood-Fill Region-Growing [1] (denoted *FloodFill*) traverses a volume in a flood-fill manner using a 3D spherical kernel of varying sizes. The seed-points and kernel dimensions are determined automatically based on image content (e.g. local gradients), while the kernel movement criteria for specific kernels are inferred from a set of training examples. This approach is composed of five stages: 1) definition of the 3D kernel; 2) determination of the movement criteria; 3) automatic seed initialisation; 4) flood-fill region growing and 5) region merging.

Performance Evaluation: The most widely adopted approach for the evaluation of segmentation techniques is the comparison of algorithm-generated segmentations to manually delineated ground-truth data using some similarity measure [69, 70]. The manual delineation of volumetric data, however, is a laborious task - particularly when the dataset is large and each data sample contains a large number of objects. Furthermore, we note several additional factors detracting from the suitability of this traditional evaluation protocol in the current context.

Firstly, we emphasise that the data used in this study was collected for an independent study [71] and ground-truth delineations were not performed during this initial data-gathering procedure. Furthermore, exhaustive records of the objects present in each of the scans were not kept, making the generation of ground-truth data impractical (and likely inaccurate).

Secondly, we stress that in the context of baggage security screening, segmentation is seen as an intermediate step in the classification and/or recognition of illicit materials within the scans. The objective of the segmentation algorithm is thus to generate sub-volumes or regions of interest for which the performance of the subsequent classifier is optimal. In related

studies [8, 20] it has been shown that optimal threat recognition is achieved on subvolumes containing single objects. In these prior studies, subvolumes were randomly cropped from whole-volume scans such that each approximately contained a single object. In order to retain some degree of contextual information, a margin of 30mm was then extended around each of the cropped objects. The importance (in terms of classification performance) of the inclusion of context via the extended margin has been demonstrated in a concurrent study [72], where classification performance increases significantly when context is included. The most important objective of the segmentation phase in the current context is thus to generate a subvolume that approximately contains a single object, the exact boundary of which is not critical. In light of this observation, the performance of a baggage-CT segmentation algorithm is best quantified by the number of objects in a given segmentation and the quality of those objects.

Considering the aforementioned limitations of the traditional reference-based evaluation techniques, we propose a novel quantitative analysis procedure which rewards scans segmented into the correct number of high-quality subvolumes, where high-quality refers to the *single-objectness* of the subvolume. The proposed evaluation technique thus relies only a set of test images containing known numbers of objects - no knowledge of the actual object boundaries is required. In particular, the overall segmentation score for a given image is computed as the average RFS of each segmented object multiplied by the error in the number of segmented objects:

$$\mathrm{RFS}_S = \underbrace{\left(\frac{1}{N_S}\sum_{i=1}^{N_S}\mathrm{RFS}_i\right)}_{\text{Average RFS}} \times \underbrace{\left|1 - \frac{|N_T - N_S|}{N_T}\right|}_{\text{Penalty Term (PT)}} \tag{12}$$

where N_T is the true number of objects in the image; N_S is the number of segmented objects and RFS_i is the quality score for the object i. Note that $PT \in [0, 1)\ \forall N_S \neq N_T$ and $PT = 1$ otherwise. All segmentations containing the incorrect number of components will thus be penalised. Grady *et al.* [12] have suggested a similar measure obtained by averaging the AQUA scores for each segmented object in a given image. This approach does not penalise cases where the incorrect numbers of objects are segmented from a given image. We addresses this shortcoming via the introduction of the penalty term in Equation 12. It should be noted that the proposed measure is likely to be biased in favour of refinement by CCA and IDT (as both are inherently controlled by it).

Finally, all performance evaluations are performed with and without metal artefact reduction [37]. A more comprehensive investigation of the impact of artefacts and artefact reduction on segmentation and classification performance within baggage-CT security screening has been conducted in a concurrent study [72]. Within this study, it is demonstrated that dedicated artefact reduction techniques have limited impact on overall performance.

4.1. Data

Various datasets were used in the experiments presented in this study. All images are represented in Modified Hounsfield Units (MHU) where the CT densities at each pixel fall in the range $[0, 60000]$ with air calibrated to 0 MHU and water calibrated to 10 000 MHU. Images are characterised by anisotropic voxel resolutions of $1.56 \times 1.61 \times 5$mm.

Quality measure evaluations: The GMM used in the AQUA model is built using a training set composed of 80 manually cropped single-object segments. In order to evaluate each of the quality measures (AQUA and random-forest based models) a separate test set containing 194 manually-cropped single-object segments and 415 manually-cropped and algorithm-generated multi-object segments has been created (e.g. Figure 4). The test samples vary in size (depending on the object(s) in the scan).

The performance of the AQUA measure is evaluated by examining the histogram of AQUA scores for each of the samples in the test set. In order for the AQUA measure to be successfully incorporated into the proposed segmentation framework, it is required to provide a good separation between the good and bad object segments in this histogram.

The performance of each of the random-forest-based scores is evaluated by a Leave-One-Out (LOO) cross-validation procedure. Note that the process is considered a binary classification task where single-object segments represent the positive class. Receiver Operating Characteristic (ROC) curves (computed by varying the discrimination threshold on the forest posterior - Equation 10), the Area Under the Curve (AUC) and the optimal (false-positive;true-positive) operating points are computed to illustrate the performance of each of the descriptor types. Finally, the histograms of RFS are generated for each method to illustrate the separation of single and multi-object segments.

Segmentation evaluations: Qualitative analysis of the four segmentation algorithms is performed using four, cluttered whole volume baggage-CT scans obtained on the Reveal CT-80DR scanner (Figure 12 (a) - (d)). Quantitative analysis of the four segmentation algorithms is performed using a set of 30 cropped baggage-CT scans obtained on the Reveal CT-80DR scanner. Each of the volumes in the set are cropped such that they contain a known number

of objects. All volumes are composed of 512×512 axial slices and the number of slices in each volume ranges from 92 to 112. The random forests used to guide the segmentation refinements in these experiments are built using a set of 80 manually-cropped single-object segments (the same set used to build the GMM) and 80 manually-cropped multi-object segments (a subset of the 415 multi-object test set mentioned above).

5. Results

Performance was evaluated on 3D volumetric baggage-CT imagery obtained on a CT-80DR dual-energy baggage scanner manufactured by Reveal Imaging Inc which produces volumes with low anisotropic resolutions of $1.56 \times 1.61 \times 5$mm. The following restrictions were placed on the segmented objects. All voxels with intensities lower than a predefined threshold of 1000 Modified Hounsfield Units (MHU) were considered to belong to the background and thus set to zero prior to segmentation. The minimum permissible object volume was set to 50 cm^3 and all objects in the final segmentations smaller than 50 cm^3 were thus discarded (set to zero).

5.1. Quality Measure Results

In order to evaluate the performance of the AQUA measure when applied to the CT-80DR data, a GMM using 9 Gaussians and a full-rank covariance matrix was fitted over the PCA-reduced feature vectors of the GMM training segments (Section 4.1). The number of PCA coefficients retained was selected such that approximately 99% of the feature variance in the training set was retained. The resulting AQUA scores for the 609 object test set (Section 4.1) are shown in Figure 6. Considerable overlap between the two object classes (single and multi-object) is evident, such that no clear separation boundary can be established. It is unlikely that this formulation of the AQUA measure will lead to satisfactory segmentations using the current dataset.

In the evaluation of the random-forest-based quality measures, the forest parameters were fixed for all feature types. The number of tests performed for each node split was set to $0.7Dim_F$ (where Dim_F is the dimensionality of the feature vector under consideration) - this value was fixed for all nodes in a given forest; trees were grown to a maximum depth of $D_T = 10$, with a lower bound of $IG_{min} = 10^{-4}$ on the information gain and forests contained 30 trees. It was found that using these settings resulted in tree growth terminating prior to maximum depth and thus no tree pruning was performed.

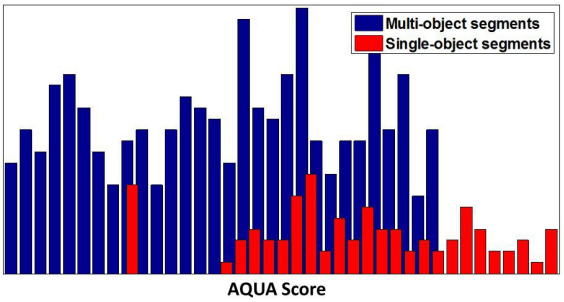

Figure 6: Separation of good and bad object segments using AQUA score [12]

.

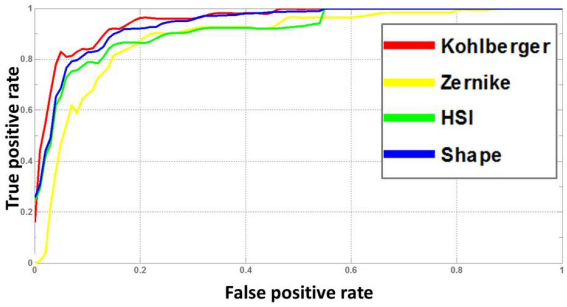

Figure 7: ROC curve for leave-one-out cross validation testing of random forest segmentation-quality measures.

Descriptor	**AUC**	**Optimal Operating Point (TPR, FPR)**
Kohlberger [39]	0.971	(0.960, 0.098)
Zernike [59]	0.863	(0.862, 0.240)
HSI [60]	0.901	(0.871, 0.160)
Shape [58]	0.942	(0.800, 0.036)

Table 1: LOO cross validation results for random forest segmentation-quality measures: Area Under (ROC) Curves (AUC) and optimal operating points.

The ROC curves and corresponding AUC and optimal operating points for the LOO evaluation of the random forest quality measures are shown in Figure 7 and Table 1 respectively. Additionally, the histogram analysis results for each descriptor type are shown in Figures 8 - 11. Based on the results and recommendations of Megherbi *et al.* [58], the 3D Zernike descriptors were generated using a maximal order of 20, yielding a 121-dimensional descriptor. The HSI was computed using a bin-width of 0.005, resulting in a 200-dimensional HSI descriptor. These settings resulted in a 321-dimensional combined 3D shape descriptor.

The Kohlberger descriptor yielded the best LOO cross-validation results (Figure 7 and Table 1) with an AUC = 0.971 and an optimal operating point on the ROC curve of (0.098, 0.960) - significantly outperforming all 3 shape-based descriptors. All four random-forest-based measures yielded superior separations of the single and multi-object segments (Figures 8 - 11) compared to the GMM-based AQUA results (Figure 6). The Kohlberger descriptor, in particular, resulted in good separation of the classes, despite the relatively high false-positive rate at its optimal operating point (Table 1). It is also worth noting that the computation of the Kohlberger descriptor [39] is considerably less computationally demanding than the Zernike [59], HSI [60] and 3D shape [58] descriptors. Based on the results of this preliminary

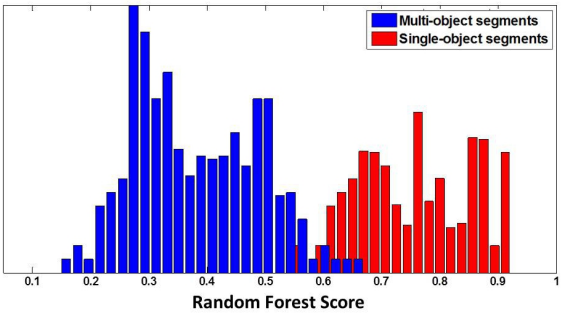

Figure 8: RFS histogram for Kohlberger descriptor [39]

.

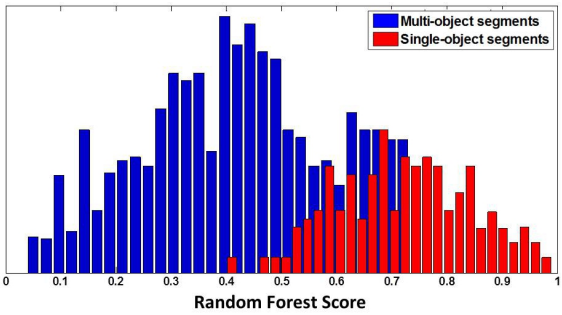

Figure 9: RFS histogram for Zernike descriptor [59]

.

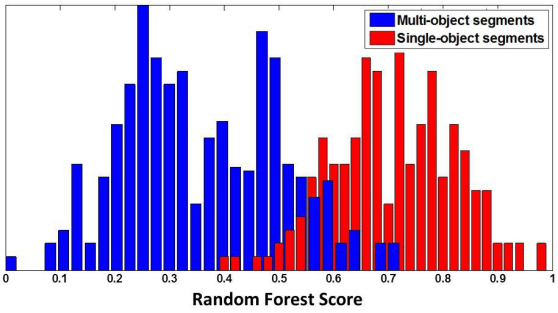

Figure 10: RFS histogram for HSI descriptor [60]

.

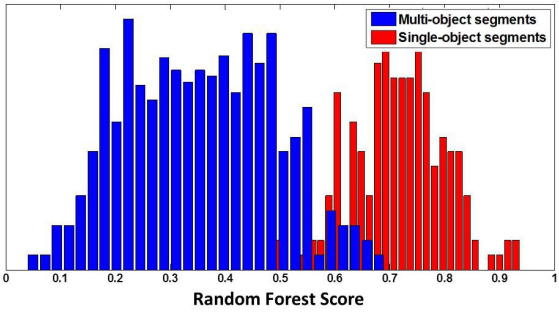

Figure 11: RFS histogram for 3D shape descriptor [58]

.

experimentation, the random forest measure using the Kohlberger descriptor was used in all subsequent evaluations. The optimal operating point for the Kohlberger descriptor occurred at a threshold of 0.73 (i.e. $(\mathbf{v} \in \mathbb{R}^{42})$ = single-object if $p(c|\mathbf{v}) > 0.73$). This threshold was used for τ_{RFS} in the segmentation-refinement procedure.

5.2. Segmentation Results

The coarse segmentations were created using $N_\tau = 10$ equally-spaced thresholds. In addition to the reasons discussed in Section 3.1, prior experimentation determined that this threshold range favoured under-segmentations (as desired) and reduced the number of objects eliminated/missed during the coarse segmentations. CCA segmentation refinement was performed using an RFS threshold of $\tau_{RFS} = 0.73$ and a hot-points threshold of $\tau_{HP} = 300$ (chosen empirically). IDT refinement [3] was implemented using a lattice-connectivity of 6 and a hot-points threshold of $\tau_{HP} = 300$. The optimal value for the quality threshold τ_γ, used as a termination criterion in the recursive application of IDT [12] was determined empirically by visually comparing candidate segmentations. FloodFill [1] was performed using

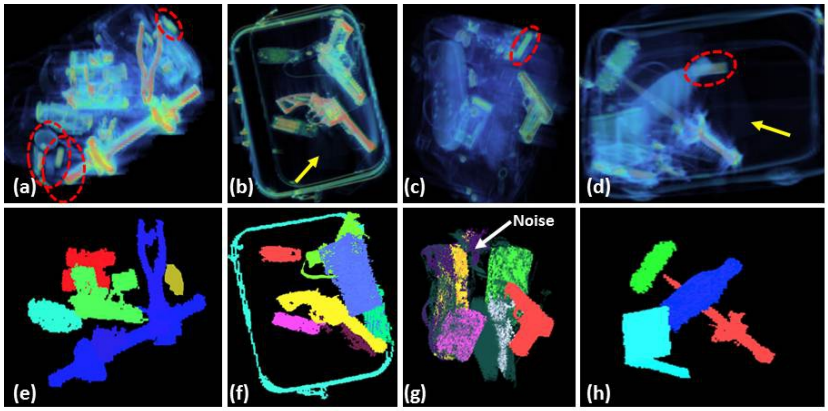

Figure 12: Segmentation test images and coarse segmentations: (a) - (d) Input scans (e) - (h) Coarse segmentations (with MAR pre-processing). Objects missed by coarse segmentation are indicated (circles and arrows).

four spherical kernels with radii $r = \{1, 2, 3, 4\}$. The movement polynomial was fitted over 70 training points obtained from 25 separate scans. SymRG was implemented using the following symmetric function:

$$g(p,q) = \begin{cases} \text{TRUE} & \text{if} |f(p) - f(q)| \leq \tau_c \\ \text{FALSE} & \text{otherwise} \end{cases} \tag{13}$$

where $f(p)$ and $f(q)$ are the intensities of voxels p and q respectively and τ_c is a user-defined constant threshold.

Figures 12 (e) - (h) show the coarse segmentations produced by the DEI thresholding process, which were used as input to the CCA and IDT segmentation-refinement procedures. Metal artefact reduction [37] was applied to the input images prior to generating the coarse segmentations. As expected, several objects are under-segmented (e.g. pliers and dumbbells in (a) and (e)). While the majority of the objects appear to have been well segmented in all four examples, two object types were commonly eliminated/missed by the coarse segmentations: 1) small cylindrical objects - encircled in red in Figure 12 and 2) thin, low-density magazines - indicated with arrows in Figure 12. The paperback book in Figure 12 (c) was well segmented in (g) - indicating that it is not the material characteristics alone of the magazines that resulted in their elimination. A more likely cause is the positioning and geometry of the magazines: in both scenarios, the magazines are lying flat against the bottom of the case/bag making them difficult to distinguish (even for the human observer) from the actual bag (on account of similar densities, their lack of bulk and noise). It is worth noting that low-density objects and thin, sheet-like objects have been known to pose difficulties for all previous baggage-segmentation

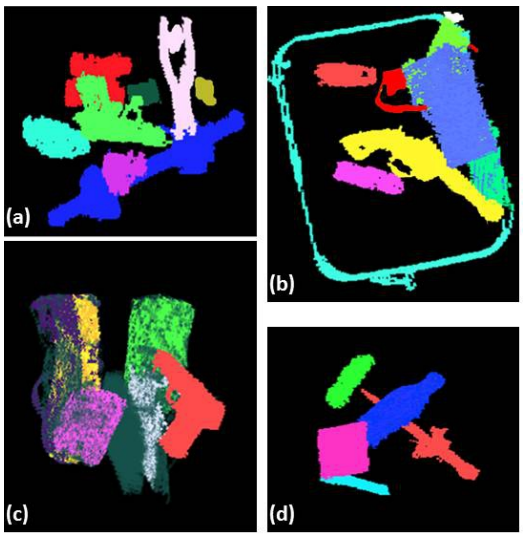

Figure 13: CCA results.

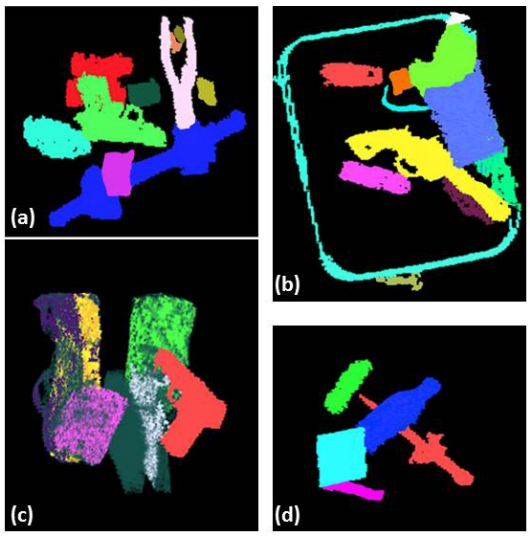

Figure 14: IDT [3] results.

approaches [1, 11, 12].

The final segmentation results produced by each of the approaches (with metal artefact reduction) are shown in Figures 13 - 17. CCA (Figure 13) and IDT (Figure 14) produced similar results as both rely on the same coarse segmentations and RFS to determine which components require refinement. The results thus differed only in those components which required refinement. In general, IDT produced superior refinements - especially evident in the test images (a) and (b). Consider, for example, the test image in Figure 13 (a) (CCA) and Figure 14 (a) (IDT), where IDT produced superior partitions in five hot-points regions. Figure 15 illustrates these regions in the original coarse segmentations and shows the computed hot-points and the post-refinement RFS. The object boundaries produced by IDT are better defined in all five cases, resulting in higher RFS (for the individual objects). Nonetheless, CCA correctly split the coarse segmentations at all hot-point regions (with the exceptions of regions 3 and 4) and produced corresponding improvements in the RFS. CCA refinements at regions 3 and 4 (Figure 15) were most likely rejected based on the resulting components not meeting the minimum permissible object size. Note that segmentation refinement, in addition to splitting merged objects, possesses denoising characteristics, as illustrated in test image (c). The coarse segmentation (Figure 12 (g)) exhibits what appears to be noise/artefacts to the right of the sole of the shoe. This noise has been removed in the corresponding regions in both the CCA (Figure 13 (c)) and IDT (Figure 14 (c)) refinements.

The segmentations produced by SymRG (Figure 16) were noticeably poorer compared to CCA and IDT. In particular, the results are characterised by under-segmentations (indicated by solid circles) and missed segmentations - where object regions have been incorrectly set as

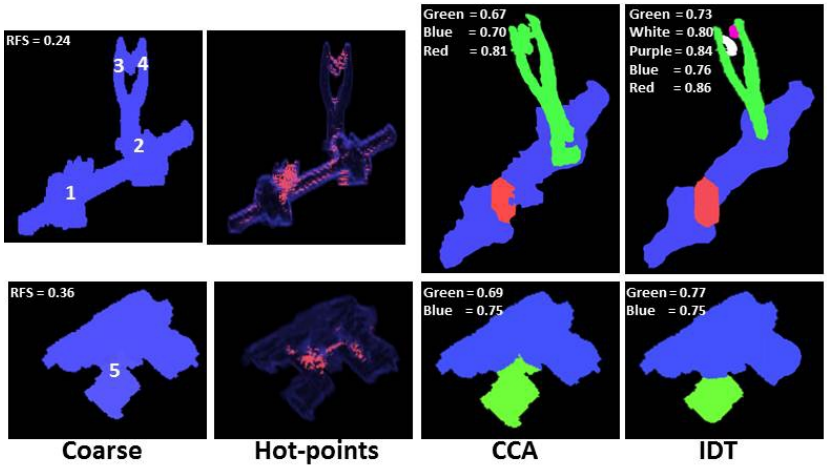

Figure 15: Segmentation refinement using CCA and IDT [12]. Hot-points in red.

background (indicated by dotted circles). This suggests the necessity for post-segmentation splitting and merging operations and explains the complexity in this regard of the segment-and-carve baggage-segmentation approach of Song *et al.* [11] (which employs a total of 5 splitting and/or merging operations). Refining the segmentations in this way significantly expands the input parameter space and hence the degree of user interaction. Accurate segmentations consequently depend on careful parameter tuning and suffer from increased computational demand. It is worth noting, however, that SymRG did capture several objects (or parts thereof) which were missed by the DEI coarse segmentations (Figures 12) - notably, the cylindrical structures in Figures 16 (a) (turquoise) and (c) (blue). Furthermore, the segmentations of the regions corresponding to the hot-point regions 3 and 4 in Figure 15, were more accurately segmented by SymRG compared to CCA (Figure 13 (a)).

Similarly to SymRG, FloodFill [1] (Figure 17) produced segmentations characterised by ill-defined object boundaries, under-segmentations (indicated by solid circles) and missed segmentations (indicated by dotted circles). The most evident shortcomings of FloodFill, however, are the poorly-defined object boundaries. This is particularly apparent for the handguns in Figures 17 (a) and (c) (compared to the equivalent CCA (Figure 13) and IDT (Figure 14) segmentations). Similarly to SymRG, the FloodFill segmentation of test image (a) captured objects which were missed by the DEI coarse segmentations (and hence the CCA and IDT results). In general, the segmentations produced by SymRG and FloodFill were of an inferior quality to the corresponding CCA and IDT segmentations.

Segmentation results without the application of metal artefact reduction are shown in Figures 18 - 22. The discriminative power of the coarse DEI segmentations (Figure 18) deteriorated significantly when metal artefact reduction was not applied. As expected, CCA

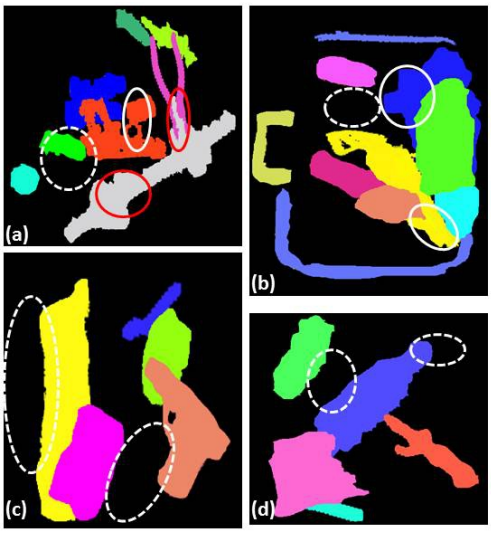

Figure 16: SymRG [2] results. Examples of under-segmentations (solid circles) and incorrect background assignments (dotted circles) indicated.

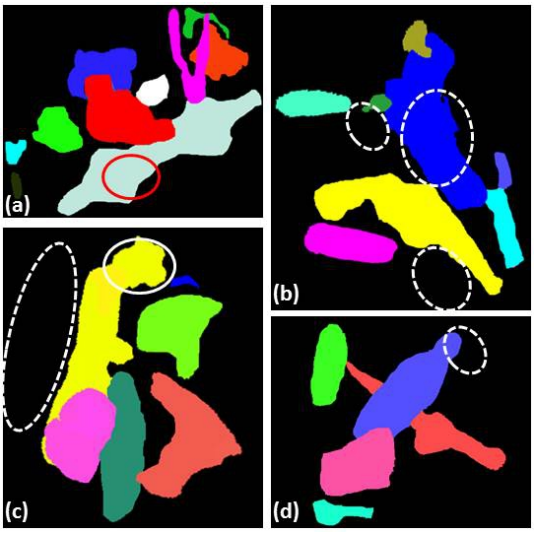

Figure 17: FloodFill [1] results. Examples of under-segmentations (solid circles) and incorrect background assignments (dotted circles) indicated.

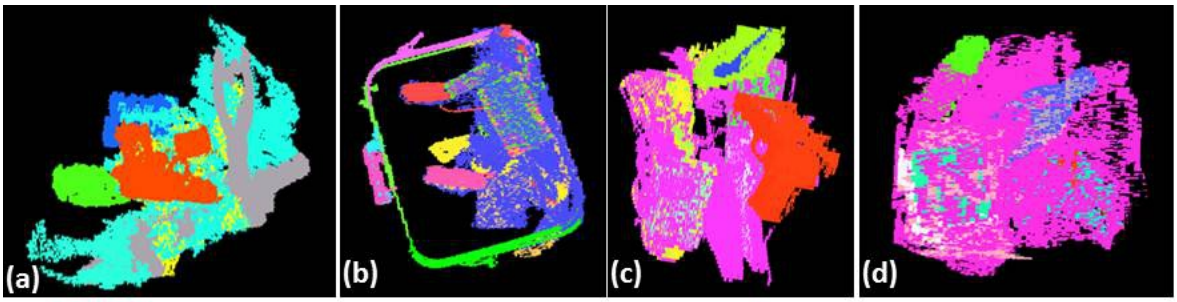

Figure 18: Coarse DEI image segmentations without MAR pre-processing.

(Figure 19) and IDT (Figure 20) segmentations suffered as a result. CCA produced segmentations characterised by a considerably higher number of under-segmentations (multiple objects labelled as a single object) and background noise (compared to the corresponding results with MAR - Figure 13). The results suggest that the high-frequency streaking artefacts result in the merging of nearby objects, making object spitting by simple connected component analysis less effective. Although IDT (Figure 20) was able to successfully split several fused objects which CCA could not (e.g. pliers and dumbbell in test image (a)), the segmentations are similarly corrupted by background noise and exhibit an increase in the number of under-segmentations.

SymRG (Figure 21) and FloodFill (Figure 22) showed a similar decline in performance in the absence of MAR. In addition to several cases of under-segmentations and missed-segmentations, similar to those produced by CCA and IDT, SymRG and FloodFill segmentations are further characterised by several examples of over-segmentations (e.g. dumbbell in

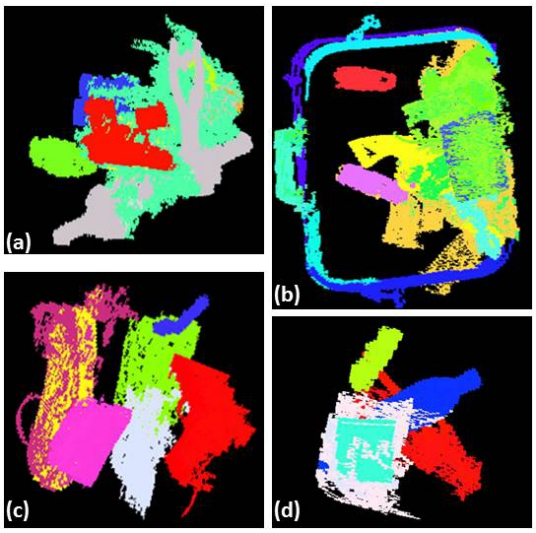

Figure 19: CCA results without MAR pre-processing.

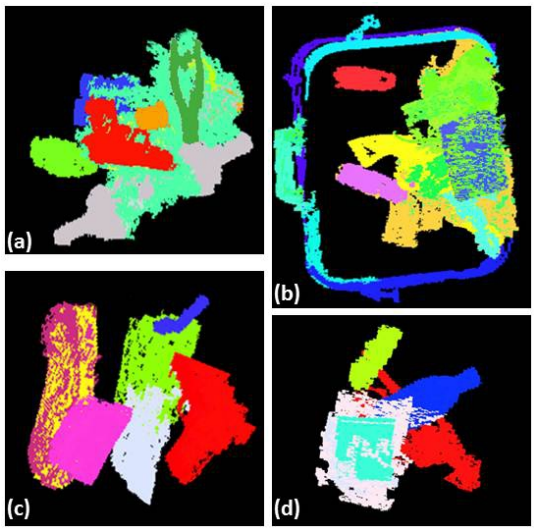

Figure 20: IDT [3] results without MAR pre-processing.

Figures 21 and 22 (a) and handgun in Figures 21 and 22 (c)).

The processing times of each of the segmentation techniques when applied to the test images with and without MAR pre-processing are shown in Figure 23. With the exception of SymRG [2], all techniques exhibited an increase in processing time when metal artefact reduction was not performed. The computational demand of SymRG is by nature dependent only on the dimensions of the input image [2] and the resulting processing times were thus consistent ($\sim 165s$) for the artefact-reduced and original volumes. CCA was the most efficient of the four techniques when operating on the artefact-reduced images, with processing times ranging from 94s to 155s. These times, however, increased by approximately 90% when MAR was not performed, making it less efficient than SymRG. The processing times of FloodFill [1] varied significantly from image-to-image (ranging from 249s to 548s for the artefact-reduced images) and appear to be largely dependent on the complexity of the image. IDT [12] was, as expected, the most computationally intensive and yielded consistently high processing times (ranging from 352s to 1238s for the artefact-reduced volumes). Both IDT and FloodFill exhibited significant increases in processing times when applied to the original volumes, with times in excess of 20 minutes. Such high processing times detract from the practical usability of these approaches (in their current states), particularly in the security-screening domain, where the demands on low processing times are paramount.

The quantitative results, with and without metal artefact reduction, are illustrated in Figures 24 - 27 and summarised in Tables 2 - 3. Figure 24 shows the total segmentation RFS for the artefact-reduced test images computed according to Equation 12. CCA and IDT [12]

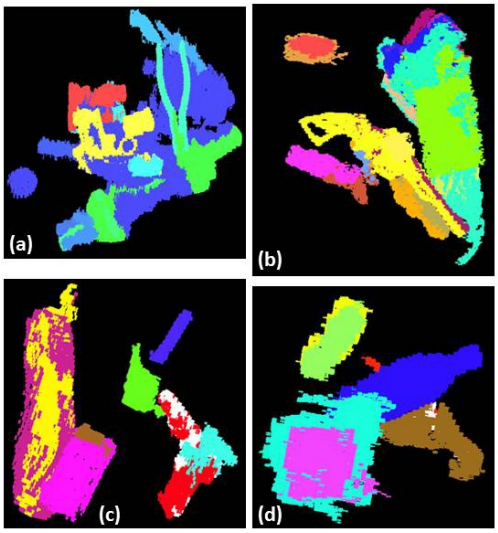

Figure 21: SymRG [2] results without MAR pre-processing.

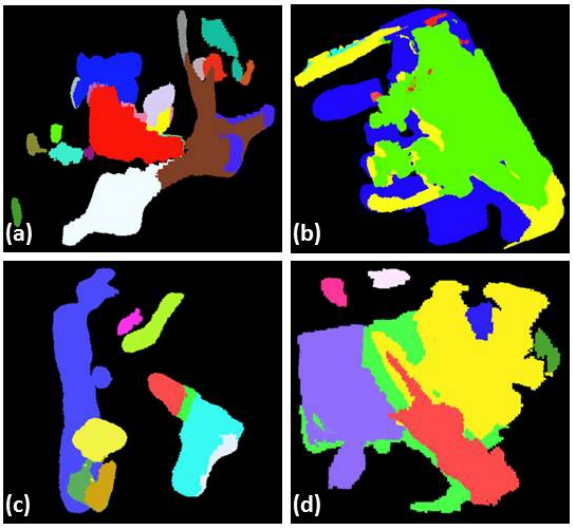

Figure 22: FloodFill [1] results without MAR pre-processing.

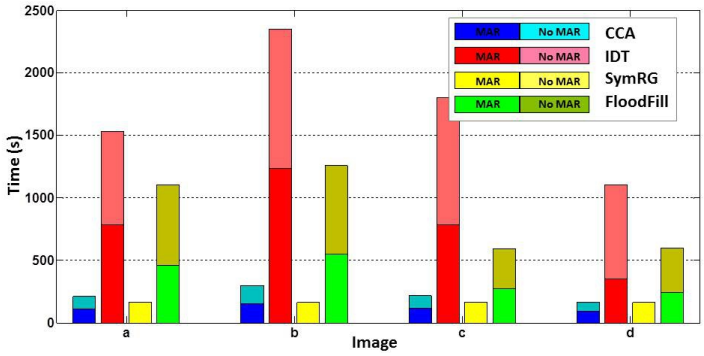

Figure 23: Segmentation processing times for test images (Figures 12 (a)-(d)) with and without MAR pre-processing.

yielded significantly higher segmentation scores compared to SymRG [2] and FloodFill [1] for all 30 test images. In particular, IDT produced on average the highest quality segmentations ($\overline{\text{RFS}}_S = 0.94$), which may be attributed to both the high quality of the individual components in each segmentation (as observed in the qualitative results) as well as the high accuracy in the number of objects segmented in each test image (an average error of only 0.1 - Table 2). Figure 25 shows that IDT segmented the correct number of objects in 27/30 images and the remaining 3 images (test images 11,23,24) each contained a discrepancy of only a single object. Although CCA produced the incorrect number of objects in 15/30 test images (Figure 25), the discrepancies were low (≤ 2). Furthermore, the mean segmentation quality remained high ($\overline{\text{RFS}}_S = 0.89$ - Table 2), indicating that the segmentation quality of the individual objects in each image was high. SymRG and FloodFill performed significantly poorer with mean scores of ($\overline{\text{RFS}}_S = 0.51$) and ($\overline{\text{RFS}}_S = 0.57$) respectively (Table 2) and the incorrect

Method	$\overline{\mathbf{RFS}}_S$	$\overline{\lvert N_T - N_S \rvert}$
CCA	0.89 ± 0.05	0.53 ± 0.57
IDT [12]	0.94 ± 0.02	0.10 ± 0.31
SymRG [2]	0.51 ± 0.08	1.73 ± 1.14
FloodFill [1]	0.57 ± 0.11	1.37 ± 1.25

Table 2: Quantitative results of four segmentation approaches with MAR: total Random Forest Score (Equation 12) and error in number of objects segmented. Results averaged over 30 volumes containing known numbers of objects.

Method	$\overline{\mathbf{RFS}}_S$	$\overline{\lvert N_T - N_S \rvert}$
CCA	0.58 ± 0.09	1.80 ± 1.13
IDT [12]	0.69 ± 0.09	1.13 ± 0.78
SymRG [2]	0.39 ± 0.17	2.93 ± 1.31
FloodFill [1]	0.41 ± 0.21	2.80 ± 2.33

Table 3: Quantitative results of four segmentation approaches without MAR pre-processing: total Random Forest Score (Equation 12) and error in number of objects segmented. Results averaged over 30 volumes containing known numbers of objects.

number of segmented objects in 26/30 images and 23/30 images respectively (Figure 25).

Figure 26 shows a decline in the segmentation quality for each method for all 30 images when metal artefact reduction was not applied. Figure 27 additionally shows that the number of over and/or under-segmented images also increased for all four methods (CCA = 27/30; IDT = 24/30; SymRG = 30/30; FloodFill = 25/30). The decline in performance was fairly consistent for all four techniques, with IDT again producing on average the highest quality segmentations ($\overline{\text{RFS}}_S = 0.69$), followed by CCA ($\overline{\text{RFS}}_S = 0.58$), FloodFill ($\overline{\text{RFS}}_S = 0.41$) and SymRG ($\overline{\text{RFS}}_S = 0.39$) (Table 3). This significant decline in performance, coupled with the increase in processing time (Figure 23), demonstrates the detrimental effects that image noise and artefacts have on the segmentation process and emphasises the importance of an effective metal-artefact-reduction process.

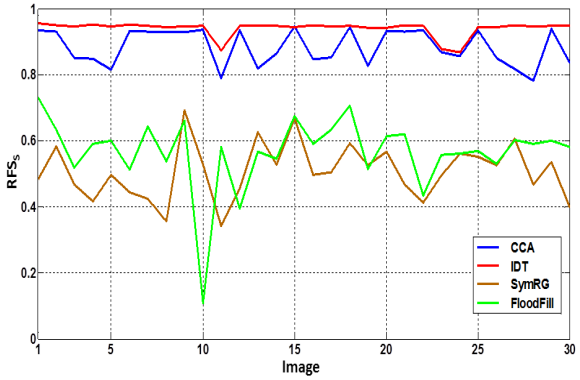

Figure 24: Overall segmentation quality scores for 30 artefact-reduced test images containing known numbers of objects.

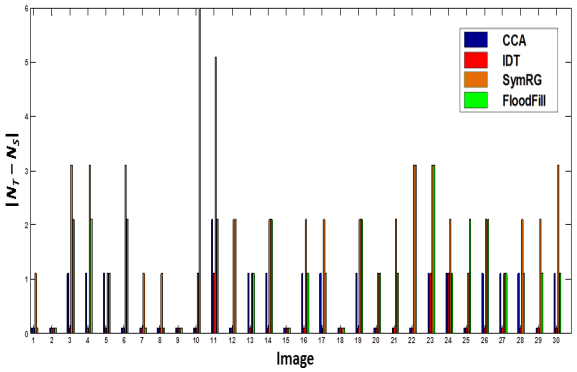

Figure 25: Errors in numbers of objects segmented for 30 artefact-reduced test images containing known numbers of objects.

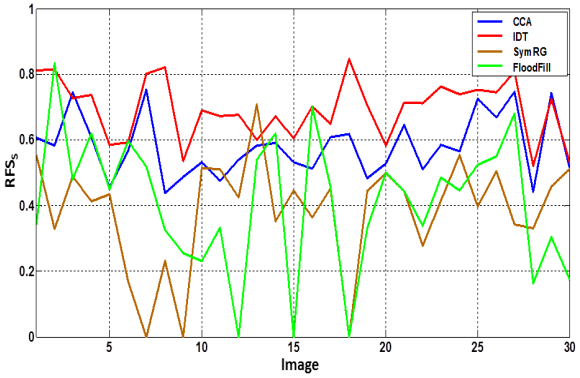

Figure 26: Overall RFS (Equation 12) without MAR pre-processing.

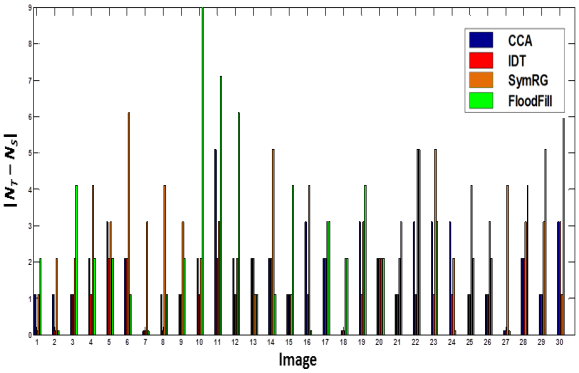

Figure 27: Errors in numbers of objects segmented for 30 test images without MAR pre-processing.

6. Conclusion

We have presented a novel materials-based technique for the 3D segmentation of unknown objects from low-resolution, dual-energy volumetric baggage-CT imagery. Within the proposed approach, four novel contributions have been made: 1) a materials-based coarse segmentation technique based on the Dual-Energy Index (DEI) [7]; 2) a random-forest-based model for measuring the quality of individual object segments; 3) a random-forest-based model for measuring the quality of entire segmentations and 4) a segmentation-refinement procedure for splitting fused objects.

Initial coarse segmentations are generated using a characterisation of the chemical composition of an image using the DEI [7], simple thresholding operations and connected component analysis. The quality of the individual components of the coarse segmentations is determined using a novel random-forest-based evaluation metric (the Random Forest Score (RFS)), which is trained to recognise high-quality, single-object segments. Segmented objects are represented

using the 42D (Kohlberger) descriptor described in [39], which is shown to outperform three 3D shape-based descriptors: 1) the 3D Zernike descriptor [59]; 2) the Histogram of Shape Index (HSI) [60] and 3) a hybrid shape descriptor [10]. Preliminary experimentation has demonstrated the superiority, in the current context, of the proposed RFS using the Kohlberger descriptor over the related Automated QUality Assessment (AQUA) measure [12]. Based on the RFS of a given coarse segmentation, low-quality individual object segments are subjected to an object-partitioning operation which splits fused objects at automatically-detected regions using a simple connected component analysis. A second novel segmentation-quality measure is presented for quantifying the quality of full segmentations based on the RFS of the constituent objects and the error in the number of objects segmented.

An experimental comparison between the proposed segmentation approach (denoted CCA) and three state-of-the-art volumetric segmentation techniques (the Isoperimetric Distance Tree (IDT) [3, 12, 67]; Symmetric Region Growing (SymRG) [2] and 3D flood-fill region growing (FloodFill) [1]) has been performed. Qualitative performance analysis, using four realistic, cluttered baggage scans, has demonstrated that IDT and CCA generate higher (visual) quality segmentations relative to SymRG and FloodFill. Although IDT is shown to outperform CCA in partitioning fused objects in the DEI-generated coarse segmentations, it is characterised by high processing times and is significantly outperformed by CCA in this regard. Low-density, sheet-like objects (e.g. magazines) are shown to pose difficulties for all four methods (an observation which has been made in the majority of related studies [1, 11, 12]).

A quantitative analysis, using the proposed Random Forest Score (RFS) for image segmentations and a set of volumes containing known numbers of objects, substantiates the observations made in the qualitative analysis. Particularly, IDT and CCA consistently outperform SymRG and FloodFill in terms of segmentation quality and in terms of segmentation accuracy (with reference to the number of objects segmented from each image). Finally, the importance of metal artefact reduction is demonstrated by the significant decline in performance for all four segmentation techniques, across all evaluation metrics considered (qualitative and quantitative), when MAR was not considered.

The observations made in this study indicate that the proposed CCA segmentation approach is well-suited to the task of volumetric image segmentation - particularly in the baggage security-screening domain, where the demands for low processing times are paramount.

Throughout this study, we have highlighted several areas for future work. Due to the

sensitivity of security data and the related challenges in data gathering, a relatively limited dataset has been used in this study (particularly in terms of the availability of ground-truth information). An important direction for future work is thus the expansion of the current dataset, including a rigorous documentation of the baggage contents, including geometric and material properties. A case-by-case fine-tuning of the DEI thresholds used for coarse segmentations (e.g. by DEI histogram analysis) is highlighted as an additional area for future work. Finally, the incorporation of the proposed segmentation approach into a fully-automated 3D object-classification technique will be evaluated.

References

[1] D. F. Wiley, D. Ghosh, C. Woodhouse, Automatic segmentation of CT scans of checked baggage, in: Proceedings of the 2nd International Meeting on Image Formation in X-ray CT, 2012, pp. 310–313.

[2] S. Y. Wan, W. E. Higgins, Symmetric region growing, IEEE Transactions on Image Processing 12 (2003) 1007–1015.

[3] L. Grady, Fast, quality, segmentation of large volumes - isoperimetric distance trees, in: Proceedings European Conference on Computer Vision, volume 3, 2006, pp. 449–462.

[4] Y. Zhou, K. Panetta, S. Agaian, 3D CT baggage image enhancement based on order statistic decomposition, in: Proceedings of the IEEE International Conference on Technologies for Homeland Security, 2010, pp. 287–291.

[5] B. R. Abidi, Y. Zheng, A. V. Gribok, M. A. Abidi, Improving weapon detection in single energy X-ray images through pseudocoloring, IEEE Transactions on Systems, Man, and Cybernetics 36 (2006) 784–796.

[6] S. Singh, Explosives detection systems (EDS) for aviation security, Signal Processing 83 (2003) 31–55.

[7] T. R. Johnson, Medical Radiology/Diagnostic Imaging: Dual Energy CT In Clinical Practice, Springer, 2011.

[8] G. Flitton, T. Breckon, N. Megherbi, A 3D Extension to Cortex Like Mechanisms for 3D Object Class Recognition, in: IEEE Proceedings International Conference on Computer Vision and Pattern Recognition, 2012, pp. 3634–3641.

[9] G. Flitton, T. P. Breckon, N. Megherbi, A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery, Pattern Recognition (2013).

[10] N. Megherbi, J. Han, T. P. Breckon, G. T. Flitton, A comparison of classification approaches for threat detection in CT based baggage screening, in: Proceedings of the IEEE International Conference on Image Processing, 2012, pp. 3109–3112.

[11] C. Crawford, H. Martz, H. Piena, Segmentation of objects from volumetric CT data, 2013. Northeastern University, Boston. This report can be found at: `https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf`.

[12] L. Grady, V. Singh, T. Kohlberger, C. Alvino, C. Bahlmann, Automatic segmentation of unknown objects, with application to baggage security, European Conference on Computer Vision, Springer, 2012, pp. 430–444.

[13] N. Megherbi, T. Breckon, G. Flitton, A. Mouton, Investigating existing medical CT segmentation techniques within automated baggage and package inspection, in: Proceedings SPIE Security and Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence, volume 8901, 2013, pp. 1–8.

[14] A. Krähenbühl, B. Kerautret, I. Debled-Rennesson, F. Mothe, F. Longuetaud, Knot segmentation in 3D CT images of wet wood, Pattern Recognition 47 (2014) 3852–3869.

[15] D. F. Wiley, Analysis of anatomic regions delineated from image data, US Patent 8,194,964 (US App. No. 12/430,545) (2009).

[16] J. A. Bangham, P. Chardaire, C. J. Pye, P. D. Ling, Multiscale nonlinear decomposition: The sieve decomposition theorem, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1996) 529–539.

[17] R. Sedgewick, Algorithms in C, 3 ed., Addison-Wesley Professional, 1998.

[18] L. G. Ugarriza, E. Saber, S. R. Vantaram, V. Amuso, M. Shaw, R. Bhaskar, Automatic image segmentation by dynamic region growth and multiresolution merging, IEEE Transactions on Image Processing 18 (2009) 2275–2288.

[19] A. Mouton, N. Megherbi, K. van Slambrouk, J. Nuyts, T. Breckon, An experimental survey of metal artefact reduction in computed tomography, Journal of X-Ray Science and Technology 21 (2013) 193–226.

[20] A. Mouton, T. Breckon, G. Flitton, 3D Object Classification in Complex Volumes using Randomised Clustering Forests, in: IEEE International Conference on Image Processing, 2014 - to appear.

[21] G. Flitton, Extending computer vision techniques to recognition problems in 3D volumetric baggage imagery [PhD Thesis], Cranfield University UK, 2012.

[22] J. Hsieh, Computed tomography: principles, design, artifacts, and recent advances, SPIE and John Wiley and Sons, 2003.

[23] R. E. Alvarez, A. Macovski, Energy-selective reconstructions in X-ray computerised tomography, Physics in Medicine and Biology 21 (1976) 733–744.

[24] W. Kalender, W. Perman, J. Vetter, E. Klotz, Evaluation of a prototype dual-energy computed tomographic apparatus. i. phantom studies, Medical Physics 13 (1986) 334–339.

[25] Z. Ying, R. Naidu, C. R. Crawford, Dual energy computed tomography for explosive detection, Journal of X-ray Science and Technology 14 (2006) 235–256.

[26] Y. Jin, Implementation and optimization of dual energy computed tomography [PhD Thesis],

University of Erlangen-Nuremberg, 2011.

[27] Z. Ying, R. Naidu, S. Simanovsky, C. R. Crawford, Method of and system for computing effective atomic number images in multi-energy computed tomography, US Patent 7,190,757 (Filed: March 13, 2007).

[28] O. Semerci, E. Miller, A parametric level-set approach to simultaneous object identification and background reconstruction for dual-energy computed tomography, IEEE Transactions on Image Processing 21 (2012) 2719–2734.

[29] K. Chuang, H. Huang, A fast dual-energy computational method using isotransmission lines and table lookup, Medical Physics 14 (1987) 186–192.

[30] W. Cai, D. Zhang, J.-G. Lee, H. Yoshida, Low-dose dual-energy electronic cleansing for fecal-tagging CT colonography, in: SPIE Medical Imaging, International Society for Optics and Photonics, 2013, pp. 86700W–86700W.

[31] A. Graser, T. R. Johnson, H. Chandarana, M. Macari, Dual energy CT: preliminary observations and potential clinical applications in the abdomen, European Radiology 19 (2009) 13–23.

[32] J. Altenbernd, T. A. Heusner, A. Ringelstein, S. C. Ladd, M. Forsting, G. Antoch, Dual-energy-CT of hypervascular liver lesions in patients with hcc: investigation of image quality and sensitivity, European Radiology 21 (2011) 738–743.

[33] A. Graser, T. R. Johnson, E. M. Hecht, C. R. Becker, C. Leidecker, M. Staehler, C. G. Stief, H. Hildebrandt, M. C. Godoy, M. E. Finn, et al., Dual-energy CT in patients suspected of having renal masses: Can virtual non-enhanced images replace true non-enhanced images?, Radiology 252 (2009) 433–440.

[34] S. Hu, E. A. Hoffman, J. M. Reinhardt, Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images, IEEE Transactions on Medical Imaging 20 (2001) 490–498.

[35] S. Ukil, J. M. Reinhardt, Smoothing lung segmentation surfaces in three-dimensional X-ray CT images using anatomic guidance, Academic Radiology 12 (2005) 1502–1511.

[36] I. Sluimer, M. Prokop, B. van Ginneken, Toward automated segmentation of the pathological lung in CT, IEEE Transactions on Medical Imaging 24 (2005) 1025–1038.

[37] A. Mouton, N. Megherbi, T. Breckon, K. Van Slambrouck, J. Nuyts, A distance driven method for metal artefact reduction in computed tomography, in: Proceedings IEEE International Conference on Image Processing, 2013, pp. 2334–2338.

[38] C. Solomon, T. Breckon, Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab, Wiley-Blackwell, 2010.

[39] T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, L. Grady, Evaluating segmentation error without ground truth, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2012, pp. 528–536.

[40] R. A. Johnson, I. Miller, J. E. Freund, Probability and statistics for engineers, Prentice-Hall, 2011.

[41] T. K. Moon, The expectation-maximization algorithm, IEEE Signal processing magazine 13

(1996) 47–60.

[42] G. McLachlan, D. Peel, Finite mixture models, Wiley. com, 2004.

[43] A. Criminisi, J. Shotton, S. Bucciarelli, Decision forests with long-range spatial context for organ localization in CT volumes, in: MICCAI Workshop on Probabilistic Models for Medical Image Analysis, 2009.

[44] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, A. Criminisi, Entangled decision forests and their application for semantic segmentation of CT images, in: Information Processing in Medical Imaging, Springer, 2011, pp. 184–196.

[45] B. Glocker, O. Pauly, E. Konukoglu, A. Criminisi, Joint classification-regression forests for spatially structured multi-object segmentation, in: European Conference on Computer Vision, Springer, 2012, pp. 870–881.

[46] G. Saiprasad, C.-I. Chang, N. Safdar, N. Saenz, E. Siegel, Adrenal gland abnormality detection using random forest classification, Journal of Digital Imaging (2013) 1–7.

[47] A. Kreshuk, C. N. Straehle, C. Sommer, U. Koethe, M. Cantoni, G. Knott, F. A. Hamprecht, Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images, PloS one 6 (2011) e24899.

[48] M. Yaqub, R. Napolitano, C. Ioannou, A. Papageorghiou, J. A. Noble, Automatic detection of local fetal brain structures in ultrasound images, in: IEEE International Symposium on Biomedical Imaging, 2012, pp. 1555–1558.

[49] V. Lempitsky, M. Verhoek, J. A. Noble, A. Blake, Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography, in: Functional Imaging and Modeling of the Heart, Springer, 2009, pp. 447–456.

[50] A. Akselrod-Ballin, M. Galun, J. M. Gomori, M. Filippi, P. Valsasina, R. Basri, A. Brandt, Automatic segmentation and classification of multiple sclerosis in multichannel MRI, IEEE Transactions on Biomedical Engineering 56 (2009) 2461–2469.

[51] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, N. Ayache, Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images, NeuroImage 57 (2011) 378–390.

[52] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. Thomas, T. Das, R. Jena, S. Price, Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2012, pp. 369–376.

[53] J. Margeta, E. Geremia, A. Criminisi, N. Ayache, Layered spatio-temporal forests for left ventricle segmentation from 4d cardiac MRI data, in: Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges, Springer, 2012, pp. 109–119.

[54] Z. Yi, A. Criminisi, J. Shotton, A. Blake, Discriminative, semantic segmentation of brain tissue in MR images, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2009, pp. 558–565.

[55] L. Monno, R. Bellotti, P. Calvini, R. Monge, G. B. Frisoni, M. Pievani, Hippocampal segmentation by random forest classification, in: IEEE International Workshop on Medical Measurements and Applications, 2011, pp. 536–539.

[56] Y. Yubin, L. Hui, Z. Qing, Content-based 3D model retrieval: a survey, Chinese journal of computers 27 (2004) 1297–1310.

[57] L. Zhang, M. J. da Fonseca, A. Ferreira, C. R. A. e Recuperação, Survey on 3D shape descriptors, Technical Report, Technical Report FCT POSC/EIA/59938/2004, 2007.

[58] N. Megherbi, G. T. Flitton, T. P. Breckon, A classifier based approach for the detection of potential threats in CT based baggage screening, in: Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 1833–1836.

[59] M. Novotni, R. Klein, Shape retrieval using 3D zernike descriptors, Computer-Aided Design 36 (2004) 1047–1062.

[60] C. Dorai, A. K. Jain, COSMOS-a representation scheme for free-form surfaces, in: Proceedings of the IEEE International Conference on Computer Vision, 1995, pp. 1024–1029.

[61] N. Canterakis, 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition, in: In 11th Scandinavian Conference on Image Analysis, Citeseer, 1999.

[62] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: Proceedings Eurographics/ACM SIGGRAPH symposium on Geometry processing, 2003, pp. 156–164.

[63] V. Venkatraman, P. Chakravarthy, D. Kihara, et al., Application of 3D Zernike descriptors to shape-based ligand similarity searching., J. Cheminformatics 1 (2009) 19.

[64] M. C. Delfour, J.-P. Zolésio, Shapes and geometries: metrics, analysis, differential calculus, and optimization, volume 22, SIAM, 2011.

[65] A. Criminisi, Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning, Foundations and Trends in Computer Graphics and Vision 7 (2011) 81–227.

[66] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.

[67] L. Grady, E. L. Schwartz, Isoperimetric graph partitioning for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 469–475.

[68] S. Y. Wan, A. P. Kiraly, E. L. Ritman, W. E. Higgins, Extraction of the hepatic vasculature in rats using 3-D micro-CT images, IEEE Transactions on Medical Imaging 19 (2000) 964–971.

[69] Y. J. Zhang, A survey on evaluation methods for image segmentation, Pattern Recognition 29 (1996) 1335–1346.

[70] W. J. Niessen, C. J. Bouma, K. L. Vincken, M. A. Viergever, Error metrics for quantitative evaluation of medical image segmentation, in: Characterization in Computer Vision, Springer, 2000, pp. 275–284.

[71] G. Flitton, T. Breckon, N. Megherbi, Object recognition using 3D SIFT in complex CT volumes, in: Proceedings British Machine Vision Conference, 2010, pp. 11.1–11.12.

[72] A. Mouton, On artefact reduction, segmentation and classification of 3D computed tomography imagery in baggage security screening [PhD Thesis], Cranfield University UK, 2014.